A proximity based macro stress testing framework

Wälchli, Boris

Abstract: In this a paper a non-linear macro stress testing methodology with focus on early warning is developed. The methodology builds on a variant of Random Forests and its proximity measures. It is embedded in a framework, in which naturally defined contagion and feedback effects transfer the impact of stressing a relatively small part of the observations on the whole dataset, allowing to estimate a stressed future state. It will be shown that contagion can be directly derived from the proximities while iterating the proximity based contagion leads to naturally defined feedback effects. Since the methodology is Random Forests based the framework can be estimated on large numbers of risk indicators up to big data dimensions, fostering the stability of the results while reducing inaccuracies in estimated stress scenarios by only stressing a small part of the observations. This procedure allows accurate forecasting of events under stress and the emergence of a potential macro crisis. The framework also estimates a set of the most influential economic indicators leading to the potential crisis, which can then be used as indications of remediation or prevention.

DOI: https://doi.org/10.1515/demo-2016-0015

Originally published at:
Wälchli, Boris (2016). A proximity based macro stress testing framework. Dependence Modeling, 4(1):251-276.
DOI: https://doi.org/10.1515/demo-2016-0015

DE GRUYTER OPEN Depend. Model. 2016; 4:251–276

Research Article **Open Access**

Boris Waelchli*

A proximity based macro stress testing framework

DOI 10.1515/demo-2016-0015
Received August 12, 2016; accepted November 2, 2016

Abstract: In this a paper a non-linear macro stress testing methodology with focus on early warning is developed. The methodology builds on a variant of Random Forests and its proximity measures. It is embedded in a framework, in which naturally defined contagion and feedback effects transfer the impact of stressing a relatively small part of the observations on the whole dataset, allowing to estimate a stressed future state. It will be shown that contagion can be directly derived from the proximities while iterating the proximity based contagion leads to naturally defined feedback effects. Since the methodology is Random Forests based the framework can be estimated on large numbers of risk indicators up to big data dimensions, fostering the stability of the results while reducing inaccuracies in estimated stress scenarios by only stressing a small part of the observations. This procedure allows accurate forecasting of events under stress and the emergence of a potential macro crisis. The framework also estimates a set of the most influential economic indicators leading to the potential crisis, which can then be used as indications of remediation or prevention.

Keywords: Random Forests, Machine Learning, Stress Testing, Early Warning Indicators, Big Data

MSC: 68T05, 68T15, 62H30

1 Introduction

Stress testing is of increasing importance in all industries. Regulatory requirements as well as renewed accounting standards are asking for macro stress tests to better safeguard against a crisis. Macro stress testing is a relatively new field. It requires testing stress-effects within the greater and most significant part of the financial system and aims at analyzing its resilience as a whole. The merits of macro stress testing are seen in the context of either crisis management or early warning indication. To manage a crisis a stress scenario is applied to known key risk indicators (KRI) and a re-mediating action is derived alike for example the determination of economic capital, whereas in an early warning indication framework the KRIs are identified themselves. In the case of early warning, scenario design is crucial ([7]). Ideally, the macro prudential scenarios should be plausible, severe and suggestive of mitigation opportunities ([10]). Apart from the obvious choice of historical scenarios, measures for plausibility of self constructed, hypothetical scenarios and algorithms and methods to find them have been suggested. However, in the shadow of the financial crisis, scholars (see for example [7]) are suggesting that scenarios might have to be implausibly severe to include the expectation of the unexpected, while especially in the case of historical scenarios scholars have formulated doubt as to whether early warning frameworks can actually work (see for example [7]). The failure of prediction ahead of the financial crisis in 2007/08 indeed casts doubt on the usage of historical data to assess the probability of an upcoming crisis.

From a methodological point of view, researchers find that most currently performed macro stress tests do not go beyond the immediate effects in the market and could be enhanced by a longer time horizon and

***Corresponding Author: Boris Waelchli:** Department of Banking and Finance, University of Zurich,Plattenstrasse 14, CH-8032 Zuerich, Switzerland, E-mail: boris.waelchli@uzh.ch

corresponding correlation/contagion and feedback effects, preferably in a nonlinear framework ([7]). Additionally it is often assumed that modeled interdependence remains stable over time while in stressed states such relations can change quickly ([7]).

At the BIS it is propagated that macro stress testing is a toolbox, not a single tool. This paper adds a tool by developing a framework that is a big data suitable model for adverse economic movements with high predictive capabilities, requiring only a few stressed inputs with the option for finding policy indicators. The proposed model is based on the proximities of a Random Forests variant on an empirical dataset. The framework does not focus on stressing individual risk indicators as is usually done but on stressing the values of all risk indicators on a subset of the observations (for example financial institutions). More specifically, a suitable sample of observations from a current state dataset today is chosen and stressed on all risk indicator's values of each of these observations. This stress should reflect the values of the variables of the stressed observation in a future state. How these observations are stressed is not the subject of this paper and not covered herein, however, it can be done taking into account common econometric models of interaction between the variables or by expert judgment. Once the chosen sample is stressed, all other not sampled observations are infected by the values of the variables of the stressed observations by means of the Random Forests proximities. This step is often called contagion. On the infected, estimated stress state data (future) a new Random Forests model is built. By iterating the contagion model, feedback effects are produced. The model thus encompasses the concept of contagion and feedback effects, inherently defined in the model. Based on the number of stress caused events in the estimated stress state data, the framework indicates whether a potential crisis could emerge due to the applied scenario. In case the framework suggests a crisis, the importance measures defined within the Random Forests algorithm allow to identify the most important variables which had the highest influence in the classification of the observations. This variables can be used to identify remediation actions against the potential crisis. Thus the proposed proximity based stress testing framework can be used as an early warning indicator as well as an instrument to identify actions to manage or prevent a crisis.

This paper shows that the resulting Random Forests model predicts future stress events accurately using a relatively small initial sample of observations, while the amount of risk indicators is theoretically not limited: Being based on Random Forests the methodology can cope with a large number of indicators being thus suitable for big data analysis and it can consequently model national and international KRIs together. The characteristics of the framework offer the advantage of robustly modeling the interdependence between observations by applying as many risk indicators as possible and by reducing the impact of estimation errors in stress scenarios. The later by using only few stressed inputs and being able to choose as such either observations which are stressed in a straight forward way or observations where the values of stressed indicators are known with a high degree of certainty. Again, it is assumed hat a model on how to stress observations is already present and the proposed framework and applied on only a few most suitable observations.

The remainder of the paper is structured as follows: the second section positions the paper in the current literature on macro stress testing and early warning frameworks under special consideration of Random Forests. In the following third section the concepts of the Random Forests variant of recursive conditional participation and the proximities is introduced. The mathematical foundation of the proposed model is laid out. The fourth section applies the model to an empirical analysis and elaborates the policy indications from the model. The final section concludes the paper.

2 Literature review

The contribution of this paper is situated in three areas of the macro stress testing literature: in general macro stress test modeling, in modeling early warning frameworks and in the application of machine learning algorithms.

Current trends in macro stress testing encompass integrating different risks, contagion- and feedback effects:

The idea of contagion is the transmission of a shock by a relatively small number of market participants (e.g. banks, sovereigns) to other or most of the other participants. To include the concept of contagion has become a common feature in macro stress testing. Some of the earlier works are by Allen and Gale [3], who model contagion effects of claims between banks; De Bandt and Hartmann [11], consider contagion effects in the broader context of systemic risk and Upper and Worms [33] specifically analyze contagion in the German interbank market.

Feedback effects describe the effects of stress and also contagion spreading between the market participants in the subsequent periods of time after the shock and contagion have occurred. Feedback are for example modeled by Jacobsen and Raszbach [23] who use an aggregate vector autoregressive model integrating several modules linking risk factors and balance sheets of corporates to show feedback from financial stability to the economy.

Elsinger at al. [15] integrate market risk, credit risk, interest rate risk and counterparty credit risk in the Austrian interbank sector. Boss et al. [8] extend the model of Eslinger et al. [15] to a three year horizon and incorporate profit risk. Considering the rules of accounting Drehmann et al. [13] create a stress test integrating credit and interest rate risk by modeling assets and liabilities simultaneously. State of the art stress tests also increasingly try to include liquidity risk (see for example [5], [34]). The currently most comprehensive model is the risk assessment model of the Bank of England ([1]), which also includes feedback effects.

Additionally, in their studies, Juselius and Kim [24] and also Drehmann et al. [12], have found that the macro econometric relationships are mostly non-linear. The BIS [7] has in its various analytical publications assessed that the focus on non-linearities and contagion/feedback effects is a priority while they doubt the potential of modeling network effects or aggregation models.

Taking up current research, the proposed model is non-linear, incorporates contagion and feedback effects and it will be shown that the stress tests performed ahead of the crisis are accurate. On the other hand, the model is integrated in the sense that the dependent variable depends on various macroeconomic factors from different areas of risk but only models their influence on each other indirectly by changes in proximities. However, since the proposed model does not specify how the stressed observations are built, another model from the literature which integrates all risk types can be applied to generate the stressed sample of observations which is used in the proposed framework.

Generally the usage of stress testing for early warning indication is not recommended by the BIS [7]. Reasons are the frequent lack of non-linearity and the usage of historical scenarios. The proposed framework is focused on early warning yet it is not built in the classic way. The early warning literature in finance mainly encompasses two approaches, firstly signaling approaches, where a threshold for specific early warning indicators is identified and secondly logit/probit approaches modeling the effects of risk indicators to identify early warning indicators. The proposed framework on the other hand estimates the number of events under stress, modeling the interaction of a large or even vast amount of indicators and then reverts back to identify the most important of the indicators for remediation of the stress effects. Additionally it is inherently non-linear. However, the initial estimation is still done on empirical data.

A recent representative of the signaling approach is Pasricha et al. [27] who apply an imbalance indicator model encompassing a large number of potential indicators. While alternatively in a recent work Babecky et al. [4] focus on developed economies and find by Bayesian model averaging that domestic housing prices, share prices, and credit growth, as well as the global variable private credit are KRIs. The proposed model on the other hand encompasses around 100 indicators and mostly developed countries and some emerging economies. Also for big data application Random Forests accepts many more indicators.

With focus on the proposed application of a Random Forests model several papers are preceding this one. The first paper is by Gosh and Gosh [18] followed by Frankel and Wei [17] who both apply decision trees on currency crises. Manasse and Roubini [25] use binary recursive trees on sovereign crises. The succeeding paper by Savona and Vezzoli [29] deals again with sovereign crises, while Duttagupta and Cashin [14] and Manasse et al. [26] study banking crises in emerging markets. Alessi and Detken [2] apply regression trees to excessive credit growth and leverage measurement. Savona and Vezzoli[29], Manasse et al. [26] as well as Alessi and Detken [2] all run some sort of Random Forests on their sample. Especially Alessi and Detken run the classic regression Random Forests by Breiman [9] to identify the most important variables to build the forest and construct a final decision tree with the important variables only. This paper applies likewise a Random Forests model but firstly not by using the classic Random Forests by Breiman but the conditional recursive partitioning forest ([21]/[31]) and not for the aim of building a final tree alike Alessi and Detken [2] but to construct a stressed state dataset (future). The most important variables are likewise identified. The conditional recursive partitioning framework is chosen above the classic Random Forests because the latter is known to be biased in the choice of splitting variables and thus in the assessment of the most important variables. The classic Random Forests prefers continuous variables to factors or discrete variables or variables with many different values to such with less values on the observations. Thus to identify the most important variables the unbiased method is preferably chosen.

Additionally the proposed model is based on a classification forest and not a regression forest. This because the classification model can be stressed and interpreted in an intuitive way: a stress situation will result in more observations being classified as events. On the other hand, regression trees and a regression forest do not produce new estimates of the dependent variable in case new data is applied, but allocate the observations to average values of the initial dependent variable. Of course, a stressed state causes more observations to be in nodes with higher predefined average values. Nonetheless is the interpretation of this outcome much more difficult especially since the highest average result was estimated before the stress situation.

3 Mathematical background

3.1 Recursive Conditional Partitioning

To define the algorithms the following dataset is introduced: Let $Y \in \{0, 1\}^m$ be m observations of the outcome of a binary event. Let $X \in \mathbb{R}^{m \times n}$ be a collection of m observations of n independent variables (risk indicators). A dataset is then denoted by $\mathcal{O} := (Y, X) \in \{0, 1\}^m \times \mathbb{R}^{m \times n}$. The m rows of the dataset $\mathcal{O}$, $\mathcal{O}_i$ will be henceforth referred to as the observations.

The conditional recursive partitioning forest and the classic Random Forests are very similar. The main difference is the splitting framework: within the conditional recursive partitioning framework the variables for splitting are selected based on maximizing the association to the dependent variable calculated by a linear statistic. Like the classic Random Forests algorithm, the nodes in each tree are split on a random sample of the total variables. Unlike the classic Random Forests, the trees are not grown on bootstrap samples but on samples without replacement. Strobl et al. [31] have shown that the bootstrap samples increase the bias in variable selection identified in the classic Random Forests. The conditional recursive partitioning framework then grows each tree in the forest in accordance with the following rules ([21]):

1. For each tree a training sample $\mathcal{TS} \subset \mathcal{O}$ of a predefined size s, s<m, $\mathcal{TS} := (Y^{\mathcal{TS}}, X^{\mathcal{TS}}) \in \{0, 1\}^s \times \mathbb{R}^{s \times n}$ is drawn.
2. At each node, test the global hypothesis of independence between $Y^{\mathcal{TS}}$ and $X^{\mathcal{TS}}$. If the hypothesis cannot be rejected, independence is assumed, the growth of the tree is stopped in the respective branch. If the hypothesis is rejected, in accordance with a predefined confidence level, the association of each independent variable with the dependent variable is tested and the variable with the highest association, as measured by the highest statistical significance (p value), is chosen as the variable to split on.
3. On the variable with the highest association, the point for the best binary split is chosen as the value which maximizes the test statistics for association. The data in the respective node is split by that value as in the classic Random Forests.
4. The steps are repeated within each tree for all trees in the forest until the global null hypothesis can no longer be rejected or another stopping criteria, alike a minimum number of observations in the respective nodes, applies.

Due to application of permutation testing by Strasser and Weber [30], where all possible permutations of the values in the learning sample are used, the following test statistics do not require knowledge of the distribution of the tested random variables:

1. First step is the general linear statistic to measure the association between $Y^{\mathcal{TS}}$ and an individual variable $X_{.j}^{\mathcal{TS}}$:
$$\mathbf{T}_j(\mathcal{TS}, \mathbf{w}) := vec\left(\sum_{i=1}^{s} w_i g_j(X_{ij}^{\mathcal{TS}}) e_2(Y_i^{\mathcal{TS}})\right)$$
The variable to split on is the $X_{.j^\star}^{\mathcal{TS}}$ with $j^\star = argmin_{j=1,\ldots,m} P_j$ and $P_j = P_{H_0^j}(c_{quad}(T_j(\mathcal{TS}, \mathbf{w}), \mu_j, \Sigma_j) \leq c_{quad}(t_j, \mu_j, \Sigma_j)|S(\mathcal{TS}, \mathbf{w}))$ with $c_{quad}(\mathbf{T}_j, \mu_j, \Sigma_j) = (\mathbf{T}_j - \mu_j)\Sigma_j^{+}(\mathbf{T}_j - \mu_j)^T$. Σ is the covariance matrix, Σ^+ is the Moore-Penrose inverse of the covariance matrix, while μ is the mean and S is the permutation of the responses as developed by Strasser and Weber [30]. Due to the application of these statistics on permutations of the samples, the statistics are conditioned on them.
In the case of classification the function g_j is the identity mapping or the zero vector with value 1 at the level k if a nominal variable with K levels is used ($e_K(k)$). The vec-operator turns a matrix by column-wise combination into a column vector.
2. If the aggregated p value of each T_j test for association cannot be rejected, thus if basically no p value is lower than a predefined level the classification tree is stopped. Hothorn et al. [20] suggest to use Bonferroni adjusted p values or minimum p values for aggregation.
3. Once the variable, $X_{.j^\star}^{\mathcal{TS}}$, with the highest association to the dependent variable is found, a similar test statistic is applied: Find best split value on the chosen variable by maximizing the test statistic over all possible subsets of the set of values:
$$A^\star = max_A c_{quad}(\mathbf{T}_{j^\star}^A, \mu_{j^\star}^A, \Sigma_{j^\star}^A)$$
with
$$\mathbf{T}_{j^\star}^A(\mathcal{TS}, \mathbf{w}) := vec\left(\sum_{i=1}^{s} w_i I(X_{ij^\star}^{\mathcal{TS}} \in A) e_2(Y_i^{\mathcal{TS}})\right)$$

Strobl et al. [31] have shown the framework to be unbiased in the choice of splitting-variables and thus the influence of specific variables can be interpreted. For further details please refer to [30] and [21].

However the focus in this paper is on the proximity measure which thus will be defined in detail, after the definition of a random forest.

Definition 1. *For a dataset of observations $\mathcal{O} \in \{0,1\}^m \times \mathbb{R}^{m \times n}$ and a single observation $\mathcal{O}_i, i \in \{1, \ldots m\}$, a random forest (prediction) is defined as*

$$RF(\mathcal{O}) \in \{0,1\}^m, \qquad RF(\mathcal{O}_i) \in \{0,1\} \tag{1}$$

assuming a binary (0,1) classification. The function behind RF is the conditional recursive partitioning framework with sampling of unique records only when building the trees.

The proximity measure is specifying a concept of distance between two observations in a Random Forests model.

Definition 2. *For two observations $\mathcal{O}_i$ and $\mathcal{O}_j$, $i, j \in \{1, .., m\}$ in a random forest,*

the proximity ρ_{ij} is defined as the share of trees in the forest where both observations are in the same terminal node. Consequently $(\rho_{ij})_{i,j} \in \mathbb{R}^{m \times m}$ is the matrix of mutual proximities between all m observations.

The proximity measure has the following characteristic:

$$\rho_{jj} = 1 \;\; \forall j \in \{1, \ldots, m\}, \qquad 0 \leq \rho_{ij} \leq 1 \;\; \forall i \neq j \in \{1, \ldots, m\} \tag{2}$$

3.2 Mathematical Derivation of the Proximity based Stress Testing Framework

The idea of proximity based stress testing is simple: the closer two observations are, the more likely they influence each other. Additionally there is no need to stress all observations but only a preferably small share of the total number of observations and then let the contagion and feedback effects do the rest.

At the end, the aim is to take a dataset of interest (current state data), choose a specific number of observations, stress those based on expert judgment or on an econometric model and apply the proposed model to construct, from the stressed sample, a stressed future dataset by contagion and feedback effects. On the stressed dataset a new Random Forests model is built resulting in the estimated stressed state of all observations and a list of the most important variables leading to it.

In detail, in the proposed framework the contagion and feedback effects are done by proximity weighted averages of the stressed inputs or, for the purpose of modeling feedback, by repeated application of proximity weighted averages. A proximity weight is simply the relative proximity between two observations i and j:

Definition 3. *Definition of proximity weights ρ^{ω}:*

$$\rho_{ij}^{\omega} = \frac{\rho_{ij}}{\sum_{k=1}^{m} \rho_{ik}} \tag{3}$$

with $i, j \in \{1, \ldots, m\}$. Consequently $\mathcal{W} := (\rho_{ij}^{\omega})_{i,j} \in \mathbb{R}^{m \times m}$ is the matrix of proximity weights. Note, $\mathcal{W}$ is not symmetric.

To derive the above described application of the framework, the following assumption must hold:

Assumption of Structural Stability - The proximities of observations evolve similarly over time. Thus for two sets of data, which are sufficiently close in time, the proximity matrices are equal. In other words, for each pair of observations an $\epsilon > 0$ exists such that:

$$(\rho_{ij}^{t})_{i,j} = (\rho_{ij}^{t+\epsilon})_{i,j} \tag{4}$$

for a specific point in time t and $i, j \in \{1, \ldots, m\}$. It can be shown that the dataset used for a Random Forests prediction can be replaced by a dataset of iterated, proximity weighted averages of the values in the very same dataset and still yield the same predictions. Further, the latter can be shown to be generated by a sample of only some observations and still yield the same predictions. It follows that, assuming structural stability, a generating sample for a future stress scenario can be used to build a stressed state dataset which, inserted in a current state Random Forests model, yields predictions of stressed events.

Proposition 1. *(Invariance of Prediction) The prediction results of a Random Forests model are equal for a dataset, $\mathcal{O} \in \{0,1\}^m \times \mathbb{R}^{m \times n}$ and the dataset of its respective proximity weighted averages: Assuming an association between Y and X on a perfectly accurate Random Forests model and proximities larger than zero between the observations of the same class and zero else, it follows that a positive integer l exists such that:*

$$RF(\mathcal{O}) = RF((Y, \mathcal{W}^l X)) \tag{5}$$

while RF(.) refers to the prediction of the forest of the events and non-events and $\mathcal{W}$ is the matrix of proximity weights.

Proof. It needs to be shown that all observations are classified as the same class before and after the application of proximity based contagion and feedback: $\mathcal{O}_i \in \{\mathcal{O}_j : RF(\mathcal{O}_j) = cl\} \Rightarrow \hat{\mathcal{O}}_i \in \{\hat{\mathcal{O}}_j : RF(\hat{\mathcal{O}}_j) = cl\}$, $cl \in \{0,1\}$ and $\hat{\mathcal{O}}_i := (Y_i, (\sum_{k=1}^{m} \mathcal{W}_{ik}^l X_{kj})_{i,.})$.

First the effects of contagion are analyzed: Because, by assumption, observations from different classes have a proximity measure of zero, it holds that the value X_{ij} of observation $\mathcal{O}_i$ on variable X_j is transformed by proximity based contagion into:

$$\hat{X}_{ij} := \sum_{k \in \{k : \mathcal{O}_k \in CL(RF(\mathcal{O}_i))\}} \mathcal{W}_{ik} X_{kj} \tag{6}$$

where $CL(RF(\mathcal{O}_i)) := \{\mathcal{O}_k : RF(\mathcal{O}_k) = RF(\mathcal{O}_i)\}$, $RF(\mathcal{O}_i) \in \{0, 1\}$. All transformed values of a variable X_j are thus weighted averages of the values of observations in the respective same class only.

The matrix of proximity weights $\mathcal{W}$ can, without restriction to generality, be written as a block diagonal matrix, sorted by the classes of the dependent variable: In the rows and columns, the observations with class one come first and second those with class zero:

$$\mathcal{W}' = \begin{pmatrix} (\rho^{\omega}_{ij})_{i,j:\mathcal{O}_i,\mathcal{O}_j \in CL(1)} & 0^{|CL(1)|\times|CL(0)|} \\ 0^{|CL(0)|\times|CL(1)|} & (\rho^{\omega}_{ij})_{i,j:\mathcal{O}_i,\mathcal{O}_j \in CL(0)} \end{pmatrix} \tag{7}$$

where $|.|$ is the cardinality of a set. The zero-block matrices, $0^{|CL(0)|\times|CL(1)|}$ and $0^{|CL(1)|\times|CL(0)|}$, result because the proximity between observations with different classes are zero. Ordering all observations (rows) in the dataset $\mathcal{O}$ in the same way as they are now ordered in the matrix of proximity weights, the estimated dataset, $\hat{\mathcal{O}}$ (excluding Y), resulting from a one off application of proximity based contagion, can be written as matrix multiplication: $\hat{X}' := \mathcal{W}'X'$ with X' being the set of independent variables X, ordered in the same order as $\mathcal{W}'$.

Second, the effects of the feedback iteration on the proximities: Iterating the estimation of the dataset means taking the proximity weighted average of the proximity weighted average iteratively. The proximity weighted average of the proximity weighted average is then $\mathcal{W}'\mathcal{W}'X'$. This results in a power sequence:

$$\mathcal{W}'X', \mathcal{W}'^2X', \ldots, \mathcal{W}'^lX'$$

$\mathcal{W}'$ is per definition a stochastic row matrix as are its non-zero diagonal blocks, defined in equation 7. Within the non-zero diagonal blocks, $(\rho^{\omega}_{ij})_{i,j:\mathcal{O}_i,\mathcal{O}_j \in CL(cl)}$, $cl \in \{0, 1\}$ the diagonal itself is non-zero and always holds the highest row value (because each observation is closest to itself). Note that it was assumed that observations within the same class have a non-zero proximity assuring that the non-zero diagonal blocks contain entries larger than 0.

Thus $\mathcal{W}'$ has no rows where all entries are zero but it has square matrices on its diagonal, with non-zero entries. Having zero entries in the upper right corner the whole proximity matrix is also of lower block-triangular form. Fulfilling this conditions Qu, Wang and Hull [28] have shown that the sequence of stochastic matrices of proximity weights, $\mathcal{W}'^k$ converges:

$$lim_{k\to\infty}\mathcal{W}'^k = \begin{pmatrix} 1_{|CL(1)|}c_1 & 0^{|CL(1)|\times|CL(0)|} \\ 0^{|CL(0)|\times|CL(1)|} & 1_{|CL(0)|}c_0 \end{pmatrix} \tag{8}$$

with c_1 and c_0 being stochastic vectors and $1_{|CL(1)|}$ being a $|CL(1)|$ times $|CL(1)|$ square matrix of ones. Since X' is stable throughout the sequence it follows that the sequence of proximity weighted averages converges likewise.

Third, the effects of the feedback iteration on the observations conclude the proof: Due to the assumption of an association between Y and X it can without restriction of generality be assumed that for each variable X_v, larger values of this variable are more often associated with class 1 and lower values more often with class 0. Then, because of the convergence of the sequence of averages, the limits of each class must be different and it must hold that for each variable X_v a positive number l_v of iterations exists, such that each weighted average in class one at this point in the sequence, is larger than any weighted average in class zero:

$$\exists l_v : min((\rho^{\omega}_{ij})^{l_v}_{i,j:\mathcal{O}_i,\mathcal{O}_j \in CL(1)} X_{i,v:\mathcal{O}_i \in CL(1)}) > max((\rho^{\omega}_{ij})^{l_v}_{i,j:\mathcal{O}_i,\mathcal{O}_j \in CL(0)} X_{i,v:\mathcal{O}_i \in CL(0)})$$

Choosing the number of total iterations l as $l := max_v l_v$ allows to perfectly distinguish the classes in Y on each variable. Because it is assumed that $\mathcal{O}$ can be perfectly predicted by a Random Forests model, then $(Y', \mathcal{W}'^lX')$ can likewise be perfectly predicted by a Random Forests model and the observations are classified in the same class as by the original Random Forests analysis, with Y' being the dependent variable Y, ordered in the same order as $\mathcal{W}'$. Since the vertical order of the observations in the datasets does not influence the prediction of the random forest the result applies likewise to $(Y, \mathcal{W}^lX)$. □

Corollary 1. *(Minimal generator) For each dataset* $\mathcal{O} \in \{0, 1\}^m \times \mathbb{R}^{m\times n}$*, a minimal generating set in* $\mathcal{O}$ *exists that, by application of proximity weighted averages, generates a dataset* $\hat{\mathcal{O}}$ *such that the same Random Forests* *predictions result as for RF($\mathcal{O}$): Assuming an association between Y and X on a perfectly accurate Random Forests model and proximities larger than zero between the observations of the same class and zero else, it follows that a minimal generating set $\mathcal{S} \subset \mathcal{O}$ exists which generates the same RF results as RF($\mathcal{O}$) using proximity based contagion and feedback:*

$$RF(\hat{\mathcal{O}}) = RF((Y, \mathcal{W}^l X^{\mathcal{S}})) = RF((Y, \mathcal{W}^l X)) = RF(\mathcal{O}) \tag{9}$$

with $X^{\mathcal{S}} := (\sum_{k:\mathcal{O}_k \in \mathcal{S}} \rho^{\omega}_{ik} X_{kj})_{i,j}$.

Proof. Following the proof of proposition 1 the matrix of proximity weights $\mathcal{W}^l$ is not changed as it is built by $RF(\mathcal{O})$.

Having drawn $\mathcal{S}$, the proximity weighted average for any value is built from those observations in $\mathcal{S}$ which are within the same class as the observation of the considered value. Obviously these changes in the dataset do not affect the convergence of the product of X with the stochastic proximity weight matrix [28].

Because an association between Y and X is assumed, it holds again that on each individual variable X_v, without loss of generality, the higher values can be attributed to class 1 while the lower ones, after a certain threshold, can be attributed to class 0. Additionally, for the sake of the argument every row $i : \mathcal{O}_i \in \mathcal{O}/\mathcal{S}$ is set to a vector of zeros. Note that $\mathcal{O}$ is considered as the set of the observations. Then the following cases conclude the proof:

- If $\mathcal{S} \subset \mathcal{O}$ but $\mathcal{S} \not\subset CL(1) \wedge \mathcal{S} \not\subset CL(0)$, then $\mathcal{S}$ contains observations of both classes. Then the multiplication of the power sequence in equation 8 and $(Y, \mathcal{W}^l X^{\mathcal{S}})$ again converge to a dataset $\hat{\mathcal{O}}$ with the same Random Forests predictions as the original dataset $\mathcal{O}$.
- If on the other hand $\mathcal{S} \subset CL(j)$, $j \in \{0, 1\}$ then the observations of one of the two classes in $\hat{\mathcal{O}}$ are zero. However, since the values on the other class are larger than zero a Random Forests model can discriminate them again perfectly.

As such the assumed perfect accuracy of the forest is preserved by the sampled transformation and the prediction remains the same as on the original forest using the whole dataset. □

Corollary 2. *Assuming an association between Y and X on a perfectly accurate Random Forests model, then corollary 1 likewise holds if the condition that 'proximities between the observations of the different classes are zero' is relaxed to 'proximities between the observations of the different classes are smaller than those between observations of the same class'.*

Proof. Following the proof of proposition 1 it needs to be shown that the stochastic matrix $\mathcal{W}^l$ still converges although it is no longer of lower block-triangular form. This is indeed still the case due to Wolfowitz [37] lemma 2. Subsequently, because of lemma 2.1 of Qu et al.[28], $\mathcal{W}^l$ converges to the matrix $c \times I$, where c is a constant and I the m×m identity matrix.

Assuming again without loss of generality, the higher values on a variable X_v can be attributed in tendency to class 1 while the lower ones can be attributed in tendency to class 0, it follows that most values attributed to a class 1 observation must be above c while most values attributed to a class 0 observation must be below. Because of the assumption that observations in the same classes are closer to each other than to observations of other classes ($\rho_{ij} > \rho_{sl}, \forall i, j, s, l : cl(\mathcal{O}_i) = cl(\mathcal{O}_j)$ and $cl(\mathcal{O}_s) \neq cl(\mathcal{O}_l)$), the proximity weighted averages of most values attributed to a class 1 observation are, during the iteration, converging to c monotonically decreasing while most values attributed to a class 0 observation are converging to c monotonically increasing.

It thus holds that there exists a finite number of iterations l such that the proximity weighed averages of the values of a specific variable within class 1 are all larger than the proximity weighted averages within class 0 of the same variable. The remaining proof follows then from corollary 1. □

Remark 1.

- *The assumption of non-zero proximities within the diagonal stochastic square matrices that are larger than proximities between observations of different classes, reflects the expectation that if a model is built and accurate, the observations of the same class are sensitive to the same risk drivers and 'closer' to each other. However, in the empirical application there will be cases where this assumption does not hold.*
- *Additionally the assumption of a perfectly accurate forest will not always hold on an empirical dataset. However, using a large amount of variables and at least 5000 trees, experience has shown that the Random Forests models exhibit an average in-sample classification error of less than 1%.*
- *If all assumptions were holding and the full proximity matrix is available, then a sample size of one observation is sufficient to conform to corollaries 1 and 2. However, since the assumptions, although they are reasonable, will not fully hold in reality, the empirical application of the methodology will exhibit a deviation to the expected theoretical results. As such additional information in the form of a larger sample will add accuracy and the actual sample size is best calibrated on historical data in a respective portfolio.*

Proposition 2. *(Stress Prediction) Assuming a time series of datasets of observations $\mathcal{O}^t$ and matrices of proximity weights $\mathcal{W}(t)$ built on these datasets, an association between $Y(t)$ and $X(t)$ on a perfectly accurate Random Forests model and that proximities between the observations of the different classes are smaller than those between observations of the same class and of those observations which change classes between time t and t+1. Assuming that the assumption of structural stability holds, it follows that a minimal generating set $\mathcal{S}(t+1) \subset \mathcal{O}(t+1)$ exists which generates the same RF results as RF($\mathcal{O}(t+1)$) using proximity based contagion and feedback with the proximity information at time t:*

$$RF(Y(t+1), \mathcal{W}^l(t)X^{\mathcal{S}}(t+1)) = RF(Y(t+1), \mathcal{W}^l(t+1)X^{\mathcal{S}}(t+1)) = RF(\mathcal{O}(t+1)) \tag{10}$$

Proof. Technically the main difference between proposition 2 and corollary 2 is an unknown number of observations which will change the class due to contagion and the iteration of proximity weighted averages (feedback effects).

Since these observations, the transition observations, form part of class cl in time t and class $\neg cl$ in time t+1 their inter-class proximities (proximities between the observations of the different classes) can reasonably be assumed to be higher than those of observations which do not change class. Following the proof of corollary 2 the proximity weighted averages converge to a value c. Assuming also and again without loss of generality, that during the iteration of the proximity weighted averages most values attributed to a class 1 observation are converging to c monotonically decreasing while most values attributed to a class 0 observation are converging to c monotonically increasing. Then, because the transition observations have higher proximities to class 1 observations than the observations which remain in class 0, they increase faster towards c compared to class 0 observations. It thus exists a number of iterations l after which the maximum proximity weighted average of the values of observations attributed to class 0 on a specific variable X_v is lower than the minimum proximity weighted average of the values of the transition variables.

Since $Y(t+1)$ is assumed to be known, proposition 2 follows directly from corollary 2 and the assumption of structural stability. □

Proposition 2 is the main result in this paper. As laid out above it can be shown that the dataset in a Random Forests prediction can be replaced by a dataset of iterated proximity weighted averages generated by a subset of stressed observations yielding the same predictions as would be derived using the full stressed data. Thus under the outlined assumptions it is sufficient to stress a small number of observations (or market participants as denominated in the literature review, while in this paper the market participants are sovereigns) in order to estimate the future stress state of a dependent variable and the whole dataset. The result can be used for stress testing and early warning and allows by the concept of importance measurement to identify the main risk drivers of future event occurrences.

To apply proposition 2, two issues have to be tackled. First, as mentioned, the minimal generating set is not specified nor how it is found: In the next section, 'Empirical Study', the observations are chosen based on their proximity to all other observations to maximize contagion and feedback effects, which suits the model best. The minimum size of the set in an environment of only partially fulfilled assumptions is empirically calibrated also in the next section.

Second, in an empirical application the dependent variable $Y(t+1)$ is with exception of the stressed observations not known. $Y(t+1)$ needs thus to be estimated alike the dependent variables $X(t+1)$. More specifically the proximity based contagion and feedback is likewise applied to Y. Since Y can be considered equivalent to just another variable, proposition 2 applies as well and the values of the estimated $\hat{Y}$ will be clearly distinguishable (the accuracy of the distinction is depending on whether the assumptions are fulfilled). Yet, the estimated values will be weighted averages between 0 and 1 and $\hat{Y}$ as such not suitable to be used as dependent variable in a classification Random Forests model. To use the weighted $\hat{Y}$ the following transformation (rounding) is applied: The values of $\hat{Y}$ which are above a certain threshold τ are attributed to class 1 and and those below to class 0. The threshold τ can be found by calibration as laid out in the next section.

4 Empirical study

The Dataset

To make full use of the capabilities of Random Forests, a large number of independent variables or risk indicators should be used. Considering the aim to show that the proximity based stress testing framework can predict or warn about future crises, the used dataset should be a time series. Therefore the public and online available data of the World Bank, "World Development Indicators & Global Development Finance" has been sourced [1].

The independent risk indicators are selected from currently applied theories on GDP growth, such as tax raising, public spending, monetary policy, the liberty of the economic environment, the workforce and its education and international trade. Indicators with more than 33% missing values are excluded. Indicators that cannot be easily compared between countries such as indicators measured in local currency or other absolute values are also not included. In numbers, 104 indicators are chosen between 1998 and 2010 (with an average of 9% missing values between 1999-2010). The large number of indicators in the model can easily be coped with by Random Forests and as Biau [6] shows, there will be no distortion from variables with no predictive power. The indicators and their descriptions are listed in the appendix. Since the Random Forests based recursive conditional partitioning does not over-fit ([9]), many more indicators could theoretically be introduced.

The 12 years of data in the sample encompasses information from Australia, Austria, Belgium, Brazil, Canada, China, Czech Republic, Denmark, Finland, France, Germany, Greece, Hong Kong SAR, Hungary, India, Indonesia, Ireland, Italy, Japan and the United States. The choice of countries to be included in the sample represent mostly the developed world including some emerging economies. The specific choice of the countries is based upon data availability and quality.

As dependent variable an indicator for financial stability was chosen: This paper is considering the changes in the number of non-performing loans per country as such. The non-performing loans (NPLs) are studied in various scientific papers. Espinoza and Prasad [16] describe NPL as key macroeconomic indicator for financial stability and investigate its feedback effects over a three year period. They especially find that financial institutions with a high NPL are very sensitive to macroeconomic stress. Likewise Vatansever and Hepsen [35] argue that NPL is an important economic performance measure and apply a regression and co-

1 Online in internet: http://data.worldbank.org/indicator

integration analysis to show a significant relationship between NPLs and a list of macroeconomic indicators. Finally Inaba et al. [22] analyze the interrelationship between the increase in non-performing loans (NPLs) and the performance of the real economy in Japan, modeling first the effect of macroeconomic variables on NPLs and then the respective feedback effect of a raise in NPLs on the economy. They find significant distorting influence of NPLs.

In this paper the NPL (number of non-performing loans as share of total loans as share of GDP) is again drawn from "World Development Indicators & Global Development Finance" [2].

Since the recursive conditional partitioning framework is used as a classification algorithm, the dependent variable has to be binary. It is common that an event based on NPL movement occurs only after a certain threshold. For example when the ratio exceeds 20% (see [27]). However, in this paper an event is not based on the level of the NPL ratio itself but on the level of change between the analyzed points in time. Independent of the level of NPL a sufficiently large change in NPL indicates a crisis.

In this paper an event is defined as rise in non-performing loans (NPL) of at least 10% annually compared to the previous year. The dependent variable Y will take the value 1 for an NPL event and 0 for otherwise. Hardy and Schmieder [19] describe in their work that the NPL rise around 10% from the typical levels one year ahead of what they call an average crisis, in comparison to 25% for a severe crisis. Also Vazquez et al. [36] macro stress tested credit risk in the Brazilian banking sector and found an increase of 3.3% in long term NPL in their GDP scenario as a stress effect. This indicates that a threshold of 10% is high enough to serve as indicator for a crisis in this analysis. The plausibility of the choice is shown by observing that during the years analyzed in figure 1, the share of events defined in such a way evolve as expected: The NPL evolution shows

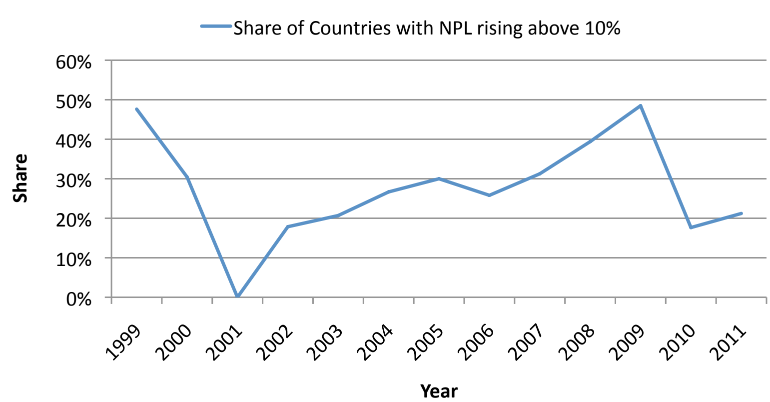

Figure 1: Share of countries in the sample with rising NPL, per year.

a macro stress period in 1999/2000 and from 2006 to 2009. This coincides with empirically observed crises. This paper claims that the model will be able to amply reproduce the share of rising NPL of each period.

Description of the Analysis Process

To assess the performance of the proposed model, a backtesting approach is applied: for a specific point in time, the succeeding two years are predicted by starting with an RF model and proximity matrix on the preceding two years. Note that a time window of two years is the minimum in this paper in order to have sufficient data points to assure the quality of the results. The results are then compared to the observed individual and macro stress events.

2 Online in internet: http://data.worldbank.org/indicator

In general, the approach would contain two estimation steps: first, the estimation of a stress state/the stressing of a subsample of the current state data and second the estimation of the full stress state in the target future using proximity based contagion and feedback effects. The aim of the paper however is to show a methodology that allows to use only a small stressed sample of the data one is interested in and construct the rest of the stress state/scenario using the proposed methodology, namely proximity based contagion and feedback. As mentioned, the stressing itself does not form part of the paper. For this reason a perfectly accurate stress scenario model is assumed by using the actually observed stress values as stress estimates: to neutralize potential errors from stress estimation. In detail: Let's define $stress_{t->t+1}(\mathcal{O})$ as the function or methodology to stress a set of observations $\mathcal{O}$ and $proxycontfed_t(\mathcal{S})$ as the application of proximity based contagion and feedb ack effects on a sample of observations $\mathcal{S}$ using the proximities in time t. Then the usual way to backtest the performance of the proposed methodology would be to compare the Random Forests (RF) classification results for $RF(\mathcal{O}(t+1))$ with $RF(proxycontfed_t(stress_{t->t+1}(Sample_t)))$, where $Sample_t$ is a sample of the data at time t and $O(t+1)$ is the observed stress state data in time $t+1$. However, this backtesting approach includes the estimation of the stress state itself, in other words, it is unclear, whether inaccuracies identified by the backtesting are due to a failure of the methodology of proximity based contagion and feedback or the chosen approach to stress the data sample.
In this paper the following approach to isolate model effects is applied: stressed sample $stress_{t->t+1}(Sample_t))$ is replaced by the actual values $S\mathcal{S}(t+1)$ of the sample in the stressed state, thus backtesting only:$RF(\mathcal{S}(t+1))withRF(proxycontfed_t(\mathcal{S}(t+1)))$. Since $\mathcal{S}(t+1)$ is a rather small subset of $\mathcal{O}(t+1)$ and since the proximities at time t are applied as proposed, the approach backtests the proposed methodology neutralizing unwanted effects from stressing data.

For example: To predict the state of the economy in 2007/2008 by proximity based contagion and feedback, a RF model is drawn on the years 2005/2006. In the next step a sample of observations from the 2007/2008 dataset is chosen and the respective observations in the 'current state' the 2005/2006 dataset are stressed by replacement by the chosen sample. Then the remaining observations in 2005/2006 are 'infected' by application of proximity based contagion: The values of the observations in each independent variable are replaced by the proximity weighted average of the respective values of the variables of the inserted observations. Note that the proximity based contagion is as described applied to the dependent variable Y^t as well while the results are again matched to the classes of Y based on the threshold τ_t for time t.

As soon as the duration of the feedback effects is one year or longer, the proximity weighted averages to update each value of each variable are calculated on the inserted stress sample as well. The iteration of the proximity based feedback and contagion on all observations is reflecting the intuition that the feedback of the effects of initially inserted stress sample observations is affecting all participants interdependently and that it is fading with time. The fading effect is a logical consequence of taking averages of averages.

On the resulting estimated stress dataset a new random forest is drawn, predicting the stressed state of the economy.

When applied to the whole time series, then the future state data is just the next period data, seen from the current state, which is not necessarily a macro stress period. As a matter of fact, in the used time series, only two financial crises are macro stress periods: The crises in 2000/2001 and the financial crisis around 2007. To consider the financial (subprime) crisis this paper will include the periods from 2006-2009. 2006 is included to consider the level of accuracy the model achieves directly before the crisis and as such to show that it can cope with a steep rise in stress events from one period to the next. However, in each period there are events in individual countries classifying as stress events based on the definition in this paper. The empirical analysis will thus assess the predictive power of the methodology in the next period of either individual stress events and of the macro stress events of the crises in 2000 and the financial crisis around 2007. The current period is considered the 'current normal' while the next period is a future state stress scenario with either individual events or a macro stress event. For clarity, the empirical data, which refers to the total empirical data used in the analysis, is composed of the current dataset (time t), which will be called current state data and the known next period, following the current state data (t+1), the stress state data in case of one of the two macro stress events or the future state data in case of individual stress events . The resulting dataset from the application of the methodology of proximity based contagion and feedback will be called estimated future state data, predicting either individual or macro stress. The drawn sample to initiate the methodology is the stress sample.

Model Accuracy

As mentioned, to assess the accuracy of the applied model a new forest is drawn on the resulting estimated future state data, including the proximity weighted update of the dependent variable Y^t. Then $\hat{Y}^t$ is predicted using this new forest and the estimated future state data. The result is compared to the empirically observed classification of Y^{t+1} and the accuracy is measured by three types of error: the type one error, the share of events which have been classified as non-events, the type two error, the share of non-events which have been classified as events and the average classification error of the two, the average error (or the average accuracy which is 1 minus the average error). In most of the Random Forests applications in the literature some form of the average error is reported.

Model Parameters and Calibration

The cforest algorithm implemented in the R 'party' package is applied ([20]), using the following parameter settings: quadratic test-statistics with splitting only variables which are associated to Y with at least 99% significance. The number of sampled variables tried at each split is set to the square root of the number of independent variables and the class weight is chosen as the inverse proportion of the number of events or non-events in the dataset, both as proposed by Breiman and Cutler [9]. For the stability of the results, 5000 trees are run for each forest.

In this study three input parameters into the Random Forests (cforest) model are calibrated: the *minimum number of observations* in a node to perform a split, the *classification threshold* τ and the *stress sample size*.

The *minimum number of observations* in a node to perform a split is calibrated to minimize the average classification error of the fitted forests. This parameter is found to be not very sensitive and set to the value two.

The classification threshold τ is calibrated for every year t such that error types one and two are as balanced as possible and as small as possible.

The *stress sample size* is calibrated to give accurate forecasts while being as small as possible. Based on the design of the proposed model those observations which are the most connected in the dataset, thus with the highest proximity measures, are best suited to cause contagion and feedback effects. Thus those observations are chosen to form the minimal generating sets, the stress sample. This is done in the following way: the observations are ordered with regard to the mode of their proximities and a predefined share is chosen from amongst the top entries of that list. The predefined share is calibrated on the empirical data as elaborated in the next section, section 4.1. The mode is taken as measure because an observation which is most often most highly connected to other observations is more contagious than an observation which has the highest mean, which could stem from a few close observations only.

The following analysis is done on a historic rolling window of 2 years (thus the analysis starts only in 1999). As implied by the theory in section 3.2 the training samples will be sampled of unique values only. The training samples have a size of 63.2% of randomly drawn data to build the trees, as proposed by Strobl et al. [31].

4.1 Backtesting Results

4.1.1 Validation and Calibration

Before starting the analysis, proposition 1 and corollary 1 are verified empirically by testing whether a Random Forests prediction can be reproduced by proximity weighted input data and a generating set. Note that feedback effects are ignored for this initial proof of concept.

Therefore, on a subset of the empirical dataset a random forest is drawn. Afterwards a sample of observations is chosen and the remaining observations are replaced by the proximity weighted average in accordance with the method outlined above. On the resulting dataset a new random forest is drawn and the prediction of the events is compared to the prediction of the events of the original random forest. Note that to verify proposition 1 and corollary 1 the generating set/stress sample is not drawn from stress/future state data but from the same current state data that the random forest is built on. The size of the sample varies from 0 to 100% to give a flavor of how large a sample should be to derive an accurate approximation to the Random Forests results on the used current state data. Figure 2 shows the evolution of the type one and type two errors in relation to the drawn sample size. The used years are 2005-07, however the results are equivalent on other

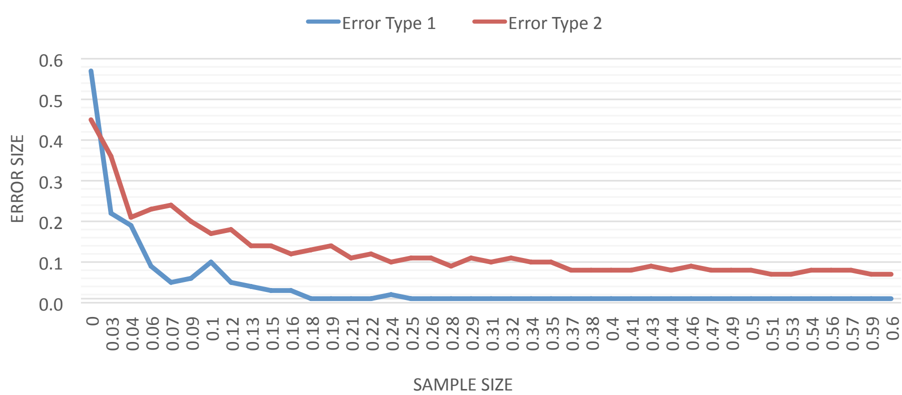

Figure 2: Type one and two errors in relation to the drawn sample initiating the proximity based contagion framework using only current state data. The errors are shown as share of the total dataset.

subsets. Using no sample and all data, class 1 has a classification error of 0 and class 0 a classification error of 6%: In other words the type two error is 0 and the type one error is 6%. As stated in corollary 1 there is a minimal set producing the same Random Forests prediction as the whole dataset which in this case is 60% of the data (in this example the stress sample is randomly drawn and not based on an analysis of the mode). Both error types however remain relatively stable until the sample is reduced to 10% and below where the errors quickly rise to around 50% which basically means that the model assigns the classes randomly. This verifies that with a fraction of a dataset (almost down to 10%) and proximity based contagion, the same results can be achieved as with the whole original dataset. Please note that it is also shown that in an environment where the assumptions made to prove corollary 1 to not fully hold, the application of proximity based contagion on a generating set does not exactly reproduce the results of the full dataset and a certain minimum amount of observations in the sample is needed to achieve a stable accuracy.

Based on this findings, the next step is a decision on how large the stress sample should be for the remainder of the analysis. Therefore the above analysis is repeated on a 2 year rolling window and with a comparison of stress sample size s of 10%, 33% and 50%. Note that again the stress sample is drawn on the respective current state data and not stress/future state data since this analysis aims at finding a suitable sample size on known data and then test the model, including the chosen sample size on unknown out-of-the-sample data. The effects on the average error of the estimated forest are depicted in table 1.

Table 1: Results of Random Forests prediction on a dataset built by proximity based contagion with full tree growing sample and 10%, 33% or 50% stress sample size. The errors are shown in % of the underlying dataset.

Years	Average-Error with 10% Sample Size	Average-Error with 33% Sample Size	Average-Error with 50% Sample Size
1999-2000	44%	35%	27%
2000-2001	49%	37%	30%
2001-2002	54%	41%	32%
2002-2003	47%	35%	26%
2003-2004	37%	26%	17%
2004-2005	43%	28%	20%
2005-2006	42%	30%	21%
2006-2007	44%	33%	22%
2007-2008	39%	27%	19%
2008-2009	38%	22%	15%
2009-2010	36%	25%	21%
Average	43%	31%	23%

The results show an expected pattern through all years of reduced average-errors whenever more data is inserted. In their similar analysis Alessi and Detken [2] derive a type one error of 38%, a type two error of 25% and thus an average-error 32%. Accordingly in this paper an average-error which is roughly below 33% is considered suitable. Based on the results in table 1 a stress sample of the size of 33% of the dataset leads on average to an average error of 31% and is thus employed throughout the paper.

4.1.2 Results

To assess the performance of the framework, the average error, the type one and the type two errors are calculated for estimated future states using proximity based contagion effects with no- and one year feedback effects. One year forecasts are the maximum forecast period considered in this paper.

Additionally, to test whether the application of the proposed proximity based contagion framework adds value at all, a random forest is drawn directly on the current state datasets, where the stress sample has been included (referred to as initial dataset) but the proximity based contagion framework has not been applied. This shows whether all the information to correctly predict a stress/future state is already included in the stress sample or added by the proximity based contagion and feedback framework.

The above described process of backtesting the performance is implemented in the following way: for an analysis at time t, 1. the data in t $\mathcal{O}(t)$ is stored as current state data and the data in $t+1$, $\mathcal{O}(t+1)$ is stored as stress/future data. 2. A random forest is drawn on the current state data and the proximity matrix is stored. 3. Using the proximity matrix the 33% of the most connected observations are identified and stressed by replacing their current state values with their stress/future state values. The current state data with the replaced values for the sampled observations is stored as 'initial dataset'. 4. The proximity based contagion and feedback effects are applied to the initial dataset which is then stored as the estimated dataset. 5. Random forests are drawn on the estimated and initial datasets separately and the respective predicted events are compared to the observed events in the stress/future state data to estimate the type one, type two and average error.

Following are the summary results (table 2) of a proximity based stress testing with a training sample of 63%, a 33% share of stressed original observations and no- and one year feedback effects on in-sample data. Modeling feedback effects is increasing the accuracy and stability of the model. The later is measured by the imbalance between type one and two errors: the lower this imbalance is and the lower its volatility is, the more stable are the results. Also the macro stress forecast for the period between 1999 to 2000 and 2006 and

Table 2: Summary of model results with no feedback and one year feedback on in-sample data. The measures are shown in % of the underlying dataset.

Performance Measure	No Feedback	1 Year Feedback
Forecast (Individual Stress)-Accuracy (1-error)	70.31%	70.39%
Macro Stress Forecast-Accuracy (1-error)	73.46%	74.3%
Average imbalance between type one& two error	16.78%	15.37%
Stdev of imbalance between type one& two error	17.14%	17.49%

2009 is more accurate than the forecast of the individual events within the full window of analysis between 1999 and 2010. However, the results shown in table 2 are derived in-sample and are thus calibrated to be most accurate. To show the accuracy and practicability of the model it has to be tested out-of-sample: the Y thresholds τ_t are calibrated in period t and applied on the estimation of the following time period t+1. The following table 3 presents the summarized results of a proximity based stress testing with a training sample of 63%, a 33% share of stressed original observations and one year feedback effects on out-of-sample data. The out-of-sample stress/future forecast is naturally less accurate than the in-sample forecast, however the

Table 3: Summary of results of in-sample and out-of-sample model application compared to the results of the unchanged dataset including the sample information. The measures are shown in % of the underlying dataset.

Performance Measure	Model IS	Model OOS	Initial Forest
Forecast (Individual Stress)-Accuracy (1-error)	70.39%	68.67%	51.88%
Macro Crisis Forecast-Accuracy (1-error)	74.3%	73.31%	54.37%
Average Imbalance between type one&two Error	15.37%	17.99%	48.24%
Stdev of Imbalance between type one&two Error	17.49%	17.53%	36.65%

decrease in accuracy is low. Compared to the forecast power of the current state data including the stress sample, the proximity based stress testing framework is significantly more accurate, reducing the average error from 48% to 31% in the case of individual stress events and from 44% to 27% in case of the macro stress events. Considering the measures on the imbalance of the type one and two errors, the model is likewise adding stability. The following table 4 shows the estimated out-of-sample errors of type one and type two as well as the errors for the forest drawn on the current state data including the stress sample. The analysis is done for a set of two year windows, the column 'years' shows the oldest year of the training dataset and the youngest of the predicted set. The detailed modeled type one and two errors are especially balanced and low in the stress state in 1999-2000 and during the crisis around 2007. Note that the accuracy in the out-of-sample testing in general and specifically on the stress states gives support to the assumption of structural stability.

However, as pointed out earlier, the assumptions on which proposition 2 is derived do not fully hold on the empirical data. This justifies to use a reduced training sample of 63% instead of a higher value or even the usage of the full sample, simply because it gives some flexibility to the model to cope with violations of the assumptions as well as inaccuracies in the stress scenario estimates in the stressed samples. This will undoubtedly occur if the later are estimated and not know as they are in the backtesting approach applied in this paper.

On the other hand, this paper propagates that one advantage of stressing only a small sample of observations is that those could be chosen to be especially easy to stress or that their risk indicator values in times of stress are known with great certainty. Thus if risk indicators of certain observations can be predicted accurately and inserted in a contagion based stress testing model, then the usage of a full instead of a 63% training sample is supposed to increase the accuracy. The next table 5 employs a 70 % as well as a full sample size

Table 4: Forecast with 63% tree growing sample, 33% stressed inputs and 1 year feedback, out-of-sample event calibration. The measures are shown in % of the underlying dataset.

Years	Average-Error Initial Dataset	Average-Error of Model	Initial Data - Error type two	Initial Data - Error type one	Model - Error type two	Model - Error type one
1999-2000	49%	26%	21%	78%	24%	29%
2000-2001	52%	37%	37%	66%	28%	47%
2001-2002	54%	45%	69%	39%	46%	44%
2002-2003	54%	32%	57%	50%	22%	42%
2003-2004	42%	28%	27%	57%	0%	55%
2004-2005	50%	31%	0%	100%	34%	28%
2005-2006	50%	27%	0%	100%	9%	45%
2006-2007	53%	28%	73%	33%	22%	33%
2007-2008	30%	23%	30%	30%	26%	20%
2008-2009	50%	30%	0%	0%	30%	30%
2009-2010	48%	38%	92%	4%	56%	20%

with a one year feedback effects. The average accuracy of the forecast of the individual stress events is indeed

Table 5: Summary of out of sample results for different sizes of the stress sample. The measures are shown in % of the underlying dataset.

Performance Measure	Model OOS - 0.632	Model OOS - 0.7	Model OOS - 0.99
Forecast (Individual Stress)-Accuracy (1-error)	68.67%	69.35%	70.41%
Macro Stress Forecast-Accuracy (1-error)	73.31%	74.33%	73.89%
Average Imbalance between type one&2 Error	17.99%	17.49%	15.82%
Stdev of Imbalance between type one&2 Error	17.53%	15.92%	13.35%

increasing with the training sample size. The macro stress forecast on the other hand is overall increasing but the accuracy using a training sample with a size of 70% is higher than the one with the full sample.

Thus basing the method on a full training sample for each tree leads indeed to higher accuracy yet obviously increases the risk of over-fitting the model. Depending on the accuracy of the model assumptions and the inserted stress data in a once built forest either the full training sample or the 63% tree growing sample size might be suitable. Note that Alessi and Detken [2] have used a training sample size of 70% in their analysis and derived a balance error of 32% for the whole time span of their analysis while this paper derives an average balance error of 30.6% for the whole sample and an average balance of 25.7% for the balance error in macro stress states.

Overall the methodology performs well and is able to predict, conditional on the accuracy of the stress sample, an upcoming future state for individual events and for the whole population. The second application of course warrants the definition of a threshold which, if sufficient observations are predicted to be in a stress state, is breached and the whole population is considered to be in a macro stress.

Analyzing figure 1 a macro stress state could be defined as a state where a third or more of the observations are individually in a stressed state. This threshold encompasses the crisis in 1999/2000 and the financial crisis around 2007. The following graph maps the observed share of events and the estimated share of events within two year windows. Note that the estimated share of events is the weighted average of the estimated events and non-events, using the overall average error as weight. The figure highlights again that the proximity based stress testing framework is able to predict macro stress states.

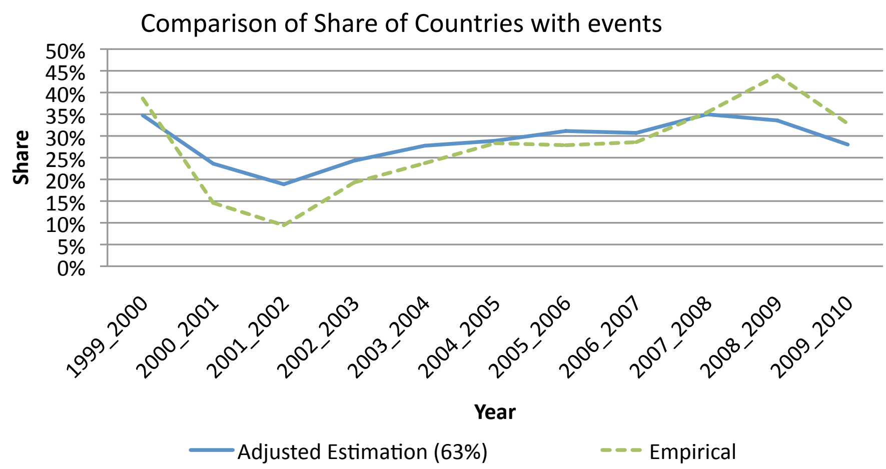

Figure 3: Comparison of share of countries in the sample.

4.2 Policy Indicators

The application of the Random Forests model allows to assess the most important variables within the proximity based stress testing framework. This adds value in the following way: The proximity based contagion and feedback model allows to identify beforehand periods of individual stress or macro stress states. Once identified, the most important variables or key risk drivers, with regard to the stress/future period, can be extracted from the Random Forests model. Thus the model shows, contingent on the correctness of the random forest, which risk drivers are important in a coming stress event and thus which risk drivers could be managed to prevent the results from the scenario. In other words the model points out policy indicators.

The implementation of the proximity based stress testing framework as a whole can thus be summarized as follows:

1. On today's data a proximity based stress testing is applied.
2. The future events are estimated.
3. If there is no significant increase in events, the dataset is not affected by the chosen stress scenario. If on the other hand there is a significant increase in events, the proximity based framework can be used to identify policy indicators to address the weaknesses in the dataset before the crisis, assumed by the stress scenario, emerges.
4. Identify the most important variables on the estimated future dataset.
5. Translate the most important variables into policy indicators by comparison of the estimated future values and the observed values today.

To be sure whether variables which are identified as important in the estimated future state data are also important in the real stress/future state data, the following backtesting is performed: the upper percentile of the distribution of the importance score of the variables in both datasets is compared and it is assessed how many variables are in both percentiles of each dataset. This shows whether the application of the above outlined methodology leads to the same importance ranking of variables in the estimated future state as in the actually observed stress/future state.

To calculate the importance score, the importance measure introduced within the recursive conditional partitioning package (party, cforest) is applied ([20]). It is defined in the following way: *Importance is defined by randomly permuting the values of a predictor variable and thus breaking its original association with the response. Thus, a reasonable measure for variable importance is the difference in prediction accuracy before and after permuting a variable, averaged over all trees* ([32]). The following table 6 points out the share of variables which are important in the estimation as well as the empirical data during the financial crisis around 2007. In a proximity based stress testing framework with a training sample size of 63%, a stress sample of the

Table 6: Persistence of most important variables in stress/future estimation and empirical data, by percentile of the importance distribution

Percentile	10%	15%	20%	25%	33%
Share of Variables which are Important in both Datasets' Percentiles: for the Macro Stress Period from 2006-2009	45%	69%	67%	62%	71%

size of 33% and 1 year feedback effects, around 45%-71% of the most important variables in the estimated future state data also contribute to the macro stress data during the stress event of the financial crisis. This is sufficient to state that important variables from the stress testing exercise actually are important in a crisis situation.

Table 7 shows in detail, which variables are actually important in both datasets. The most important shared

Table 7: List of most important variables within the stress/future state data and their occurrence in the estimated future state data. The variables are sorted in order of their importance score from most important to least important. The boundaries of the percentiles are indicated on the left.

Percentile	Variable Short Name	Variable Description	Occurrence
	GC.TAX.TOTL.GD.ZS	Tax revenue (% of GDP)	1
	TX.VAL.MRCH.R5.ZS	Merchandise exports to developing economies in South Asia (% of total merchandise exports)	1
	FM.LBL.MQMY.IR.ZS	Money and quasi money (M2) to total reserves ratio	0
	GC.TAX.GSRV.VA.ZS	Taxes on goods and services (% value added of industry and services)	1
	GC.XPN.TOTL.GD.ZS	Expense (% of GDP)	1
	FD.RES.LIQU.AS.ZS	Bank liquid reserves to bank assets ratio (%)	1
	NE.CON.GOVT.ZS	General government final consumption expenditure (% of GDP)	1
	SE.XPD.TOTL.GD.ZS	Public spending on education, total (% of GDP)	1
	NY.GNS.ICTR.ZS	Gross savings (% of GDP)	1
0.1	TX.VAL.MRCH.HI.ZS	Merchandise exports to high-income economies (% of total merchandise exports)	1
	NY.GNS.ICTR.GN.ZS	Gross savings (% of GNI)	1
	MS.MIL.XPND.ZS	Military expenditure (% of central government expenditure)	0
	NE.CON.GOVT.KD.ZG	General government final consumption expenditure (annual % growth)	1
	SH.XPD.TOTL.ZS	Health expenditure, total (% of GDP)	1
0.15	SH.XPD.PUBL	Health expenditure, public (% of total health expenditure)	0
	PA.NUS.ATLS	DEC alternative conversion factor (LCU per USD)	0
	SH.XPD.PUBL.GX.ZS	Health expenditure, public (% of government expenditure)	1
	NE.CON.TETC.ZS	Final consumption expenditure, etc. (% of GDP)	1
0.2	NY.GDS.TOTL.ZS	Gross domestic savings (% of GDP)	1
	NE.CON.PRVT.PC.KD	Household final consumption expenditure per capita (constant 2000 USD)	1
	FP.CPI.TOTL	Consumer price index (2005 = 100)	1
	NE.DAB.TOTL.ZS	Gross national expenditure (% of GDP)	1
	TX.VAL.MRCH.R3.ZS	Merchandise exports to developing economies in Latin America and the Caribbean (% of total merchandise exports)	0
0.25	SH.XPD.PUBL.ZS	Health expenditure, public (% of GDP)	1
	NE.RSB.GNFS.ZS	External balance on goods and services (% of GDP)	0
	SE.PRE.ENRR	School enrollment, preprimary (% gross)	0
	NY.GDP.DEFL.KD.ZG	Inflation, GDP deflator (annual %)	1
	TX.VAL.TECH.MF.ZS	High-technology exports (% of manufactured exports)	1
	NE.IMP.GNFS.KD.ZG	Imports of goods and services (annual % growth)	0
	TM.VAL.MRCH.AL.ZS	Merchandise imports from economies in the Arab World (% of total merchandise imports)	1
	NE.EXP.GNFS.KD.ZG	Exports of goods and services (annual % growth)	0
0.33	ST.INT.DPRT	International tourism, number of departures	0

indicators are Tax revenue, exports to South East Asia and Money and quasi money (M2) to total reserves ratio.

Indicators like money and quasi money (M2) to total reserves ratio or tax revenue can be called direct policy indicators. The reserve ratio can be changed by central banks and has an effect for example on money supply and the interest rate. Likewise if taxation is identified as an influential indicator it can be used directly as a policy indicator. On the other hand, many of the indicators, alike exports, are not direct policy indicators since they are harder to influence. However, measures can be taken to curb exports.

To translate the most important variables into actual policy indicators the average estimated levels of each indicator can be compared to the average observed current state values. Based on whether the respective indicators are direct or indirect the instruments are chosen to keep the indicator from reaching the level identified in the stress/future state. Thus for example the policy indicator of Tax revenue can be translated into the changing of taxes or the creation of new taxes. Note that a Random Forests methodology does not allow to derive exact thresholds but only indicates which variables on which levels contributed to the results and thus the stress/future state. The full extent of how policy indicators are translated into actions is not part of this paper.

5 Conclusion

In this a paper a non-linear macro stress testing methodology, the proximity based stressed testing framework, with focus on early warning and crisis remediation was developed. The development was done based on heuristic derivation and mathematical proofs. The proposed methodology builds on a conditional recursive partitioning forest: by application of its proximity measures, the effects of a small stressed sample are expanded to the whole dataset. Feedback effects are simulated by iterating the process.

Due to the inherited characteristics of Random Forests the model is compatible with the application of big data, thus allowing to use as much variables as possible to estimate the interdependence between observations or market participants as robustly as possible. While then the application of stress scenarios on only a few observations reduces the effects of inaccuracies in the scenarios as well as the possibility to use observations where either the stress/future state is easily estimated or known with great certainty.

It was shown that a Random Forests model on the estimated future state data predicts a potential crisis very well for an individual observation as well as for macro stress states by accurately forecasting the number of stress events. Likewise it has been shown that the most important variables leading to this events can be identified and potentially used as input to manage or prevent crises.

In comparison to the initial dataset and the similar model of Alessi and Detken (2014) the proposed model achieved lower average- and type one errors (table 8).

Table 8: Comparison of the results of the proposed proximity based stress testing framework against a suitable benchmark and against the dataset with only the sample included and no application of the framework (see section 4.1.2). The Errors are shown in % of the underlying dataset.

	Time Window	Average-Error	Error type one	Error type two
OOS Proximity based stress testing Framework	1999-2010%	30%	33%	26%
(33% stress sample, full training sample, 1Y feedback)	2000 2006-2008%	26%	31%	21%
OOS Proximity based stress testing Framework	1999-2010%	31%	36%	27%
(33% stress sample, 63% training sample, 1Y feedback)	2000 2006-2008%	27%	31%	22%
Initial Dataset	1999-2010%	50%	55%	45%
(sample only)	2000 2006-2008%	49%	50%	48%
Benchmark Alessi and Detken [2]	1970-2013	32%	38%	25%

Especially during the years of the crises the proximity based stress testing framework exhibits a low average classification error and similar type one and 2 errors.

The proposed proximity based stress testing framework is designed to consider most requirements formulated by the BIS ([7]) such as being non-linear, containing naturally defined contagion and feedback effects and the capability to incorporate national and international KRIs. However, initially the framework still relies on historical data. With regard to the BIS critics towards the application of early warning systems, the proposed framework addresses this by an alternative modeling of the early warning indicator: The number of the modeled stress events itself is the early warning indicator and the most important risk drivers to estimate the early warning indicator can be used to re-mediate the crisis.

Acknowledgement: I would like to thank the referees for their constructive feedback, corrections and suggestions which greatly improved the manuscript.

References

[1] Aikman, D., P. Alessandri, B. Eklund, P. Gai, S. Kapadia, E. Martin, N. Mora, G. Stern, and G. Willison (2009). Funding liquidity risk in a quantitative model of systemic stability. *Bank of England Working Paper 372*.
[2] Alessi, L. and C. Detken (2014). Identifying excessive credit growth and leverage. *ECB Working Paper Series 1723*.
[3] Allen, F. and D. Gale (2000). Financial contagion. *J. Polit. Econ. 108*(1), 1-33.
[4] Babecky, J., T. Havranek, J. Mateju, M. Rusnak, K. Smidkova, and B. Vasicek (2012). Leading indicators of crisis incidence, evidence from developed countries. *ECB Working Paper Series 1486*.
[5] Barnhill, T. and L. Schumacher (2011). Modeling correlated systemic liquidity and solvency risk in a financial environment with incomplete information. *IMF Working Paper WP/11/263*.
[6] Biau, G. (2012). Analysis of a random forests model. *J. Mach. Learn. Res. 13*(1), 1063-1095.
[7] Borio, C., M. Drehmann, and K. Tsatsaronis (2012). Stress-testing macro stress testing: does it live up to expectations? *BIS working papers 369*. Available at http://www.bis.org/publ/work369.htm.
[8] Boss, M., G. Fenz, G. Krenn, J. Pann, C. Puhr, T. Scheiber, S. W. Schmitz, M. Schneider, and E. Ubl (2008). Stress tests for the Austrian FSAP Update 2007: Methodology, scenarios and results. *OENB Financial Stability Report 15*.
[9] Breiman, L. and A. Cutler (2005). Random forests. Available at http://www.stat.berkeley.edu/users/breiman/ RandomForests/cc_home.htm.
[10] Breuer, T. and G. Krenn (1999). Durchführung von Krisentests. *Leitfaden zum Marktrisiko Band 5*. Available at http://www.oenb.at/en/img/band5dv40_tcm16-11166.pdf.
[11] De Bandt, O. and P. Hartmann (2001). *Financial Crisis, Contagion and the Lender of Last Resort*, Chapter Systemic Risk: a Survey, pp. 249–98. Oxford University Press.
[12] Drehmann, M., A. Patton, and S. Sorensen (2007). Non-linearities and stress testing. In *Risk Measurement and Systemic Risk, Proceedings of the Fourth Joint Central Bank Research Conference*, pp. 283–308. ECB, Frankfurt am Main.
[13] Drehmann, M., S. Sorensen, and M. Stringa (2010). The integrated impact of credit and interest rate risk on banks: A dynamic framework and stress testing application. *J. Bank. Financ. 34*(4), 713–729.
[14] Duttagupta, R. and P. Cashin (2011). Anatomy of banking crises in developing and emerging market countries. *J. Int. Money Financ. 30*(2), 354-376.
[15] Elsinger, H., A. Lehar, and M. Summer (2006). Risk assessment for banking systems. *Manag. Sci. 52*(9), 1301-41.
[16] Espinoza, R. and A. Prasad (2010). Nonperforming loans in the GCC banking system and their macroeconomic effects. *IMF Working Paper 10/224*. Available at https://www.imf.org/external/pubs/ft/wp/2010/wp10224.pdf.
[17] Frankel, J. A. and S. Wei (2004). Managing macroeconomic crises. *NBER Working Papers 10907*.
[18] Ghosh, S. R. and A. Ghosh (2002). Structural vulnerabilities and currency crises. *IMF Working Paper 02/9*.
[19] Hardy, D. C. and C. Schmieder (2013). Rules of thumb for bank solvency testing. *IMF Working Paper 13/232*. Available at https://www.imf.org/external/pubs/ft/wp/2013/wp13232.pdf.
[20] Hothorn, T., K. Hornik, and A. Zeileis (2006a). party: A laboratory for recursive partytioning. *Available at http://cran.r-project.org/web/packages/party/vignettes/party.pdf*.
[21] Hothorn, T., K. Hornik, and A. Zeileis (2006b). Unbiased recursive partitioning: A conditional inference framework. *J. Comp. Graph. Stat. 15*(3), 651-674.
[22] Inaba, N., T. Kozu, T. Sekine, and T. Nagahata (2005). Non-performing loans and the real economy: Japan's experience. *BIS Papers 22*.
[23] Jacobsen, T., L. J., and K. Roszbach (2005). Exploring interactions between real activity and the financial stance. *J. Financ. Stab. 1*(3), 308–341.

[24] Juselius, M. and M. Kim (2011). Sustainable financial obligations and crisis cycles. *Helsinki Economic Centre of Research Discussion Papers 313*.

[25] Manasse, P. and N. Roubini (2009). Rules of thumb for sovereign debt crises. *J. Int. Econ. 78*(2), 192–205.

[26] Manasse, P., R. Savona, and M. Vezzoli (2013). Rules of thumb for banking crises in emerging markets. *Universitá di Bologna DSE Working Paper 872*.

[27] Pasricha, G., T. Roberts, C. I., and B. Howell (2013). Assessing financial system vulnerabilities: an early warning approach. *Bank of Canada Review Autumn 2013*, 10–19. Available at http://www.bankofcanada.ca/wp-content/uploads/2013/11/boc-review-autumn13-pasricha.pdf.

[28] Qu, Z., J. Wang, and R. A. Hull (2005). Products of row stochastic matrices and their applications to cooperative control for autonomous mobile robots. *Proceedings of the American Control Conference July 2005, Portland, Oregon*.

[29] Savona, R. and M. Vezzoli (2015). Fitting and forecasting sovereign defaults using multiple risk signals. *Oxf. Bull. Econ. Stat., forthcoming*. Available at http://dx.doi.org/10.1111/obes.12052.

[30] Strasser, H. and C. Weber (1999). On the asymptotic theory of permutation statistics. *Report Series SFB 27*. Available at http://epub.wu.ac.at/102/.

[31] Strobl, C., A. Boulesteix, A. Zeileis, and T. Hothorn (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics 8:25*.

[32] Strobl, C., J. Malley, and G. Tutz (2009). An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychol. Methods 14*(4), 323–348.

[33] Upper, C. and A. Worms (2004). Estimating bilateral exposures in the german interbank market: is there a danger of contagion. *Eur. Econ. Rev. 48*(4), 827–849.

[34] Van den End, W. (2008). Liquidity stress-tester: A macro model for stress-testing banks' liquidity risk. *DNB Working Paper 175*.

[35] Vatansever, M. and A. Hepsen (2013). Determining impacts on non-performing loan ratio in turkey. *JFIA 2(4)*, 119–129.

[36] Vazquez, F., B. M. Tabak, and M. Souto (2012). A macro stress test model of credit risk for the Brazilian banking sector. *J. Financ. Stab. 8*(2), 69-83.

[37] Wolfowitz, J. (1963). Products of indecoposable, aperiodic, stochastic matrices. *Proc. Amer. Math. Soc. 14*, 733-737.

A List of applied independent risk indicators

Table 9: List of Applied Risk Indicators 1/4

Variable Name	Variable Description	Theory Class
IC.REG.COST.PC.ZS	Cost of business start-up procedures (% of GNI per capita)	Economic Environment
PA.NUS.ATLS	DEC alternative conversion factor (LCU per US$)	Monetary
NE.IMP.GNFS.KD.ZG	Imports of goods and services (annual % growth)	International Trade
NE.GDI.TOTL.KD.ZG	Gross capital formation (annual % growth)	Spending
NE.GDI.TOTL.ZS	Gross capital formation (% of GDP)	Spending
NE.EXP.GNFS.KD.ZG	Exports of goods and services (annual % growth)	International Trade
FP.CPI.TOTL	Consumer price index (2005 = 100)	Macroeconomic
FP.CPI.TOTL.ZG	Inflation, consumer prices (annual %)	Macroeconomic
NE.CON.TETC.KD.ZG	Final consumption expenditure, etc. (annual % growth)	Spending
NE.GDI.FTOT.ZS	Gross fixed capital formation (% of GDP)	Spending
IC.ISV.DURS	Time to resolve insolvency (years)	Economic Environment
NY.GDP.DEFL.ZS	GDP deflator (base year varies by country)	Macroeconomic
NY.GDP.DEFL.KD.ZG	Inflation, GDP deflator (annual %)	Macroeconomic
IC.LGL.PROC	Procedures to enforce a contract (number)	Economic Environment
TM.VAL.FOOD.ZS.UN	Food imports (% of merchandise imports)	International Trade
TM.VAL.MRCH.XD.WD	Import value index (2000 = 100)	International Trade
NE.CON.GOVT.ZS	General government final consumption expenditure (% of GDP)	Spending
NY.GNS.ICTR.ZS	Gross savings (% of GDP)	Economic Environment
NY.GNS.ICTR.GN.ZS	Gross savings (% of GNI)	Economic Environment
FM.LBL.MQMY.GD.ZS	Money and quasi money (M2) as % of GDP	Monetary
TX.VAL.MRCH.XD.WD	Export value index (2000 = 100)	International Trade
IT.NET.USER.P2	Internet users (per 100 people)	Economic Environment
SL.EMP.1524.SP.MA.ZS	Employment to population ratio, ages 15-24, male (%)	Workforce and Education
TM.QTY.MRCH.XD.WD	Import volume index (2000 = 100)	International Trade
SH.XPD.PCAP.PP.KD	Health expenditure per capita, PPP (constant 2005 international $)	Spending
GB.XPD.RSDV.GD.ZS	Research and development expenditure (% of GDP)	Spending
IP.JRN.ARTC.SC	Scientific and technical journal articles	Spending
FM.LBL.MQMY.ZG	Money and quasi money growth (annual %)	Monetary
GC.TAX.TOTL.GD.ZS	Tax revenue (% of GDP)	Tax
BX.TRF.PWKR.DT.GD.ZS	Workers' remittances and compensation of employees, received (% of GDP)	Economic Environment
GC.XPN.TOTL.GD.ZS	Expense (% of GDP)	Spending
SL.EMP.TOTL.SP.MA.ZS	Employment to population ratio, 15+, male (%)	Workforce and Education
NE.GDI.TOTL.CD	Gross capital formation (current US$)	Spending
NE.DAB.TOTL.ZS	Gross national expenditure (% of GDP)	Spending
IT.CEL.SETS.P2	Mobile cellular subscriptions (per 100 people)	Economic Environment
NE.RSB.GNFS.ZS	External balance on goods and services (% of GDP)	International Trade
NY.GNS.ICTR.CD	Gross savings (current US$)	Economic Environment
TX.VAL.TECH.CD	High-technology exports (current US$)	International Trade
NY.GDS.TOTL.CD	Gross domestic savings (current US$)	Economic Environment
BX.KLT.DINV.CD.WD	Foreign direct investment, net inflows (BoP, current US$)	International Trade

Table 10: List of Applied Risk Indicators 2/4

Variable Name	Variable Description	Theory Class
IC.REG.DURS	Time required to start a business (days)	Economic Environment
NY.GDS.TOTL.ZS	Gross domestic savings (% of GDP)	Spending
IC.LGL.DURS	Time required to enforce a contract (days)	Economic Environment
NE.GDI.FTOT.CD	Gross fixed capital formation (current US$)	Spending
BM.KLT.DINV.GD.ZS	Foreign direct investment, net outflows (% of GDP)	International Trade
IC.REG.PROC	Start-up procedures to register a business (number)	Economic Environment
IS.AIR.GOOD.MT.K1	Air transport, freight (million ton-km)	Economic Environment
TX.VAL.MRCH.WL.CD	Merchandise exports by the reporting economy (current US$)	International Trade
SH.XPD.PCAP	Health expenditure per capita (current US$)	Spending
GC.XPN.TRFT.ZS	Subsidies and other transfers (% of expense)	Spending
SL.EMP.1524.SP.ZS	Employment to population ratio, ages 15-24, total (%)	Workforce and Education
SH.XPD.PUBL.ZS	Health expenditure, public (% of GDP)	Spending
TM.VAL.MANF.ZS.UN	Manufactures imports (% of merchandise imports)	International Trade
NY.TAX.NIND.CD	Net taxes on products (current US$)	Tax
TX.VAL.MRCH.CD.WT	Merchandise exports (current US$)	International Trade
TM.VAL.MRCH.CD.WT	Merchandise imports (current US$)	International Trade
BN.CAB.XOKA.GD.ZS	Current account balance (% of GDP)	International Trade
GC.REV.SOCL.ZS	Social contributions (% of revenue)	Tax
GC.XPN.COMP.ZS	Compensation of employees (% of expense)	Economic Environment
NE.GDI.TOTL.KD	Gross capital formation (constant 2000 US$)	Spending
NE.IMP.GNFS.CD	Imports of goods and services (current US$)	International Trade
NE.GDI.FTOT.KD	Gross fixed capital formation (constant 2000 US$)	Spending
GC.XPN.INTP.ZS	Interest payments (% of expense)	Spending
BX.KLT.DINV.WD.GD.ZS	Foreign direct investment, net inflows (% of GDP)	International Trade
SH.XPD.PUBL.GX.ZS	Health expenditure, public (% of government expenditure)	Spending
BM.GSR.TRAN.ZS	Transport services (% of service imports, BoP)	International Trade
NE.EXP.GNFS.CD	Exports of goods and services (current US$)	International Trade
NE.CON.PETC.ZS	Household final consumption expenditure, etc. (% of GDP)	Economic Environment
SL.EMP.TOTL.SP.ZS	Employment to population ratio, 15+, total (%)	Workforce and Education
SL.EMP.1524.SP.FE.ZS	Employment to population ratio, ages 15-24, female (%)	Workforce and Education
TX.VAL.MRCH.R5.ZS	Merchandise exports to developing economies in South Asia (% of total merchandise exports)	International Trade
NE.RSB.GNFS.CD	External balance on goods and services (current US$)	International Trade
SE.PRM.ENRL.FE.ZS	Primary education, pupils (% female)	Workforce and Education
TM.VAL.TRAN.ZS.WT	Transport services (% of commercial service imports)	International Trade
GC.TAX.OTHR.RV.ZS	Other taxes (% of revenue)	Tax
SH.XPD.TOTL.ZS	Health expenditure, total (% of GDP)	Spending
BN.GSR.MRCH.CD	Net trade in goods (BoP, current US$)	International Trade
TX.QTY.MRCH.XD.WD	Export volume index (2000 = 100)	International Trade
TX.VAL.MANF.ZS.UN	Manufactures exports (% of merchandise exports)	International Trade
NE.CON.TETC.ZS	Final consumption expenditure, etc. (% of GDP)	Spending
SE.PRM.AGES	Primary school starting age (years)	Workforce and Education
IP.TMK.TOTL	Trademark applications, total	Economic Environment
TT.PRI.MRCH.XD.WD	Net barter terms of trade index (2000 = 100)	International Trade
NY.GSR.NFCY.CD	Net income from abroad (current US$)	International Trade
SL.TLF.CACT.MA.ZS	Labor participation rate, male (% of male population ages 15+)	Workforce and Education
NE.CON.PRVT.PC.KD	Household final consumption expenditure per capita (constant 2000 US$)	Economic Environment
FS.AST.DOMS.GD.ZS	Domestic credit provided by banking sector (% of GDP)	Economic Environment
ST.INT.TRNR.CD	International tourism, receipts for passenger transport items (current US$)	International Trade
TM.VAL.OTHR.ZS.WT	Computer, communications and other services (% of commercial service imports)	International Trade
EG.ELC.LOSS.KH	Electric power transmission and distribution losses (kWh)	Economic Environment
TX.VAL.MRCH.R3.ZS	Merchandise exports to developing economies in Latin America and the Caribbean (% of total merchandise exports)	International Trade
TM.VAL.MMTL.ZS.UN	Ores and metals imports (% of merchandise imports)	International Trade
SE.XPD.TOTL.GD.ZS	Public spending on education, total (% of GDP)	Spending
TX.VAL.MRCH.R6.ZS	Merchandise exports to developing economies in Sub-Saharan Africa (% of total merchandise exports)	International Trade

Table 11: List of Applied Risk Indicators 3/4

Variable Name	Variable Description	Theory Class
TX.VAL.MRCH.RS.ZS	Merchandise exports by the reporting economy, residual (% of total merchandise exports)	International Trade
ST.INT.TVLX.CD	International tourism, expenditures for travel items (current US$)	International Trade
SE.ENR.PRIM.FM.ZS	Ratio of female to male primary enrollment (%)	Workforce and Education
NE.EXP.GNFS.ZS	Exports of goods and services (% of GDP)	International Trade
BN.GSR.GNFS.CD	Net trade in goods and services (BoP, current US$)	International Trade
ST.INT.XPND.MP.ZS	International tourism, expenditures (% of total imports)	International Trade
TX.VAL.MRCH.HI.ZS	Merchandise exports to high-income economies (% of total merchandise exports)	International Trade
TX.VAL.FOOD.ZS.UN	Food exports (% of merchandise exports)	International Trade
SE.SEC.ENRL.GC.FE.ZS	Secondary education, general pupils (% female)	Workforce and Education
TX.VAL.TECH.MF.ZS	High-technology exports (% of manufactured exports)	International Trade
NE.CON.GOVT.KD.ZG	General government final consumption expenditure (annual % growth)	Spending
BN.KLT.DINV.CD	Foreign direct investment, net (BoP, current US$)	International Trade
SE.PRM.ENRR.FE	School enrollment, primary, female (% gross)	Workforce and Education
NE.IMP.GNFS.ZS	Imports of goods and services (% of GDP)	International Trade
SH.XPD.PRIV.ZS	Health expenditure, private (% of GDP)	Spending
BM.GSR.FCTY.CD	Income payments (BoP, current US$)	Economic Environment
FS.AST.DOMO.GD.ZS	Claims on other sectors of the domestic economy (% of GDP)	Economic Environment
BN.TRF.KOGT.CD	Net capital account (BoP, current US$)	International Trade
NY.TRF.NCTR.CD	Net current transfers from abroad (current US$)	International Trade
NE.CON.PRVT.CD	Household final consumption expenditure (current US$)	Economic Environment
BX.GSR.FCTY.CD	Income receipts (BoP, current US$)	Economic Environment
NE.CON.TETC.CD	Final consumption expenditure, etc. (current US$)	Spending
BX.GSR.TRVL.ZS	Travel services (% of service exports, BoP)	International Trade
BM.TRF.PWKR.CD.DT	Workers' remittances and compensation of employees, paid (current US$)	Economic Environment
NE.CON.PETC.CD	Household final consumption expenditure, etc. (current US$)	Economic Environment
SE.SEC.ENRL.FE.ZS	Secondary education, pupils (% female)	Workforce and Education
IT.MLT.MAIN.P2	Telephone lines (per 100 people)	Economic Environment
NE.CON.PRVT.PP.KD	Household final consumption expenditure, PPP (constant 2005 international $)	Economic Environment
NE.EXP.GNFS.KD	Exports of goods and services (constant 2000 US$)	International Trade
SL.TLF.CACT.FE.ZS	Labor participation rate, female (% of female population ages 15+)	Workforce and Education
TX.VAL.TRAN.ZS.WT	Transport services (% of commercial service exports)	International Trade
ST.INT.XPND.CD	International tourism, expenditures (current US$)	International Trade
NE.CON.TOTL.CD	Final consumption expenditure (current US$)	Spending
NE.DAB.TOTL.KD	Gross national expenditure (constant 2000 US$)	Spending
TM.VAL.INSF.ZS.WT	Insurance and financial services (% of commercial service imports)	International Trade
BN.CAB.XOKA.CD	Current account balance (BoP, current US$)	International Trade
TX.VAL.MRCH.OR.ZS	Merchandise exports to developing economies outside region (% of total merchandise exports)	International Trade
NE.IMP.GNFS.KD	Imports of goods and services (constant 2000 US$)	International Trade
FS.AST.CGOV.GD.ZS	Claims on central government, etc. (% GDP)	Economic Environment
GC.TAX.YPKG.ZS	Taxes on income, profits and capital gains (% of total taxes)	Tax
SE.PRM.ENRL	Primary education, pupils	Workforce and Education
IT.CEL.SETS	Mobile cellular subscriptions	Economic Environment
TX.VAL.TRVL.ZS.WT	Travel services (% of commercial service exports)	International Trade
IS.AIR.DPRT	Air transport, registered carrier departures worldwide	Economic Environment
NE.DAB.TOTL.CD	Gross national expenditure (current US$)	Spending
ST.INT.TRNX.CD	International tourism, expenditures for passenger transport items (current US$)	International Trade
BX.GSR.TOTL.CD	Exports of goods, services and income (BoP, current US$)	International Trade
SH.XPD.PUBL	Health expenditure, public (% of total health expenditure)	Spending
GC.XPN.OTHR.ZS	Other expense (% of expense)	Spending
NE.CON.GOVT.CD	General government final consumption expenditure (current US$)	Spending
BX.TRF.PWKR.CD.DT	Workers' remittances and compensation of employees, received (current US$)	Economic Environment
SE.SEC.AGES	Secondary school starting age (years)	Workforce and Education
SL.TLF.TOTL.FE.ZS	Labor force, female (% of total labor force)	Workforce and Education
BX.GSR.GNFS.CD	Exports of goods and services (BoP, current US$)	International Trade
NE.TRD.GNFS.ZS	Trade (% of GDP)	International Trade
TG.VAL.TOTL.GD.ZS	Merchandise trade (% of GDP)	International Trade

Table 12: List of Applied Risk Indicators 4/4

Variable Name	Variable Description	Theory Class
NE.CON.PETC.KD	Household final consumption expenditure, etc. (constant 2000 US$)	Spending
TX.VAL.MRCH.AL.ZS	Merchandise exports to economies in the Arab World (% of total merchandise exports)	International Trade
TX.VAL.MRCH.R4.ZS	Merchandise exports to developing economies in Middle East and North Africa (% of total merchandise exports)	International Trade
BX.TRF.CURR.CD	Current transfers, receipts (BoP, current US$)	International Trade
ST.INT.TVLR.CD	International tourism, receipts for travel items (current US$)	International Trade
TX.VAL.MMTL.ZS.UN	Ores and metals exports (% of merchandise exports)	International Trade
ST.INT.DPRT	International tourism, number of departures	International Trade
TM.VAL.SERV.CD.WT	Commercial service imports (current US$)	International Trade
BN.TRF.CURR.CD	Net current transfers (BoP, current US$)	International Trade
SE.SEC.ENRL.GC	Secondary education, general pupils	Workforce and Education
ST.INT.RCPT.CD	International tourism, receipts (current US$)	International Trade
TX.VAL.OTHR.ZS.WT	Computer, communications and other services (% of commercial service exports)	International Trade
GC.TAX.GSRV.RV.ZS	Taxes on goods and services (% of revenue)	Tax
BM.GSR.NFSV.CD	Service imports (BoP, current US$)	International Trade
NE.CON.PRVT.KD	Household final consumption expenditure (constant 2000 US$)	Spending
FS.AST.PRVT.GD.ZS	Domestic credit to private sector (% of GDP)	Economic Environment
BM.GSR.GNFS.CD	Imports of goods and services (BoP, current US$)	International Trade
IT.MLT.MAIN	Telephone lines	Economic Environment
BM.TRF.PRVT.CD	Private current transfers, payments (BoP, current US$)	International Trade
BX.PEF.TOTL.CD.WD	Portfolio equity, net inflows (BoP, current US$)	International Trade
SL.TLF.CACT.ZS	Labor participation rate, total (% of total population ages 15+)	Workforce and Education
BX.GSR.MRCH.CD	Goods exports (BoP, current US$)	International Trade
MS.MIL.XPND.GD.ZS	Military expenditure (% of GDP)	Spending
BM.GSR.TRVL.ZS	Travel services (% of service imports, BoP)	International Trade
GC.XPN.GSRV.ZS	Goods and services expense (% of expense)	Spending
SL.EMP.TOTL.SP.FE.ZS	Employment to population ratio, 15+, female (%)	Workforce and Education
BX.GSR.NFSV.CD	Service exports (BoP, current US$)	International Trade
ST.INT.RCPT.XP.ZS	International tourism, receipts (% of total exports)	International Trade
SE.PRM.ENRR	School enrollment, primary (% gross)	Workforce and Education
SL.TLF.TOTL.IN	Labor force, total	Workforce and Education
TX.VAL.SERV.CD.WT	Commercial service exports (current US$)	International Trade
NE.CON.TETC.KD	Final consumption expenditure, etc. (constant 2000 US$)	Spending
BX.GSR.TRAN.ZS	Transport services (% of service exports, BoP)	International Trade
BM.GSR.MRCH.CD	Goods imports (BoP, current US$)	International Trade
NE.CON.GOVT.KD	General government final consumption expenditure (constant 2000 US$)	Spending
BG.GSR.NFSV.GD.ZS	Trade in services (% of GDP)	International Trade
BM.GSR.TOTL.CD	Imports of goods, services and income (BoP, current US$)	International Trade
BX.GSR.CMCP.ZS	Communications, computer, etc. (% of service exports, BoP)	International Trade
TM.VAL.MRCH.AL.ZS	Merchandise imports from economies in the Arab World (% of total merchandise imports)	International Trade
ST.INT.ARVL	International tourism, number of arrivals	International Trade
SE.PRM.ENRR.MA	School enrollment, primary, male (% gross)	Workforce and Education
TM.VAL.TRVL.ZS.WT	Travel services (% of commercial service imports)	International Trade
NE.CON.PRVT.PP.CD	Household final consumption expenditure, PPP (current international $)	Economic Environment
SE.SEC.DURS	Secondary education, duration (years)	Workforce and Education
BN.GSR.FCTY.CD	Net income (BoP, current US$)	Economic Environment
SE.PRM.DURS	Primary education, duration (years)	Workforce and Education
BN.KAC.EOMS.CD	Net errors and omissions, adjusted (BoP, current US$)	International Trade